

18.10.2004

日本国特許庁  
JAPAN PATENT OFFICE

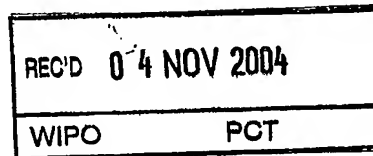
別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日  
Date of Application: 2003年11月17日

出願番号  
Application Number: 特願2003-386702  
[ST. 10/C]: [JP 2003-386702]

出願人  
Applicant(s): 日本電気株式会社

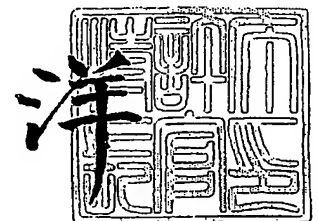


PRIORITY DOCUMENT  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH  
RULE 17.1(a) OR (b)

2004年 9月29日

特許庁長官  
Commissioner,  
Japan Patent Office

小川



BEST AVAILABLE COPY

【書類名】 特許願  
【整理番号】 34103834  
【提出日】 平成15年11月17日  
【あて先】 特許庁長官 殿  
【国際特許分類】 G06F 17/60  
【発明者】  
    【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内  
    【氏名】 襲田 勉  
【発明者】  
    【住所又は居所】 東京都港区芝五丁目 7 番 1 号 日本電気株式会社内  
    【氏名】 山下 慶子  
【特許出願人】  
    【識別番号】 000004237  
    【氏名又は名称】 日本電気株式会社  
【代理人】  
    【識別番号】 100123788  
    【弁理士】  
    【氏名又は名称】 宮崎 昭夫  
    【電話番号】 03-3585-1882  
【選任した代理人】  
    【識別番号】 100088328  
    【弁理士】  
    【氏名又は名称】 金田 暢之  
【選任した代理人】  
    【識別番号】 100106297  
    【弁理士】  
    【氏名又は名称】 伊藤 克博  
【選任した代理人】  
    【識別番号】 100106138  
    【弁理士】  
    【氏名又は名称】 石橋 政幸  
【手数料の表示】  
    【予納台帳番号】 201087  
    【納付金額】 21,000円  
【提出物件の目録】  
    【物件名】 特許請求の範囲 1  
    【物件名】 明細書 1  
    【物件名】 図面 1  
    【物件名】 要約書 1  
    【包括委任状番号】 0304683

**【書類名】 特許請求の範囲****【請求項 1】**

ラベル値が既知のデータを既知データとしラベル値が未知のデータを未知データとして前記既知データの集合と前記未知データの集合とを格納する記憶装置と、複数の学習機械とを使用する能動学習方法であって、

前記複数の学習機械が、前記既知データに関し、前記記憶装置からそれぞれ独立にサンプリングを行った後に学習を行う段階と、

前記学習の結果として、前記複数の学習機械の出力結果を統合して出力する段階と、

前記複数の学習機械が、前記記憶装置から未知データを取り出して予測を行う段階と、

前記予測の結果に基づいて次に学習すべきデータを計算して出力する段階と、

前記次に学習すべきデータに対応するラベル値を入力する段階と、

前記ラベル値が入力されたデータを前記未知データの集合から削除して前記既知データの集合に追加する段階と、

を有し、

前記既知データをサンプリングするとき、前記複数の学習機械による学習の結果を統合するとき、及び、前記複数の学習機械による予測から次に学習すべきデータを計算するとき、のうちの少なくとも1つにおいて、均等でない重み付けを実行する、能動学習方法。

**【請求項 2】**

前記既知データ及び前記未知データにおいてグループ分けを行う段階をさらに有し、前記次に学習すべきデータを計算する際に、グループで散らばるようにデータを選択する、請求項 1 に記載の能動学習方法。

**【請求項 3】**

前記既知データ及び前記未知データにおいてグループ分けを行う段階をさらに有し、前記予測を行う段階において、グループで散らばるように前記各学習機械に前記未知データを供給する、請求項 1 に記載の能動学習方法。

**【請求項 4】**

ラベル値が既知のデータを既知データとしラベル値が未知のデータを未知データとして前記既知データの集合と前記未知データの集合とを格納する記憶装置と、

既知データの学習及び未知データの予測を行う複数の学習機械と、

前記学習機械ごとに設けられ、前記記憶装置から前記既知データをサンプリングして対応する学習機械に入力する複数のサンプリング装置と、

前記各学習機械が前記既知データに基づいて行った学習の結果を統合する第 1 の統合手段と、

前記各学習機械が前記未知データに基づいて行った予測の結果から次に学習すべきデータを計算して出力する第 2 の統合手段と、

前記次に学習すべきデータに対応するラベル値を入力する結果入力手段と、

前記ラベル値が入力されたデータを前記未知データの集合から削除して前記既知データの集合に追加する制御手段と、

前記サンプリング装置ごとにサンプリング時の重みを設定するサンプリング重み付け手段と、

を有する能動学習システム。

**【請求項 5】**

ラベル値が既知のデータを既知データとしラベル値が未知のデータを未知データとして前記既知データの集合と前記未知データの集合とを格納する記憶装置と、

既知データの学習及び未知データの予測を行う複数の学習機械と、

前記学習機械ごとに設けられ、前記記憶装置から前記既知データをサンプリングして対応する学習機械に入力する複数のサンプリング装置と、

前記各学習機械が前記既知データに基づいて行った学習の結果を統合する第 1 の統合手段と、

前記各学習機械が前記未知データに基づいて行った予測の結果から次に学習すべきデータ

タを計算して出力する第2の統合手段と、

前記次に学習すべきデータに対応するラベル値を入力する結果入力手段と、

前記ラベル値が入力されたデータを前記未知データの集合から削除して前記既知データの集合に追加する制御手段と、

前記第1の統合手段で前記学習の結果を統合する際に用いられる重みを設定する予測重み付け手段と、

を有する、能動学習システム。

【請求項6】

前記第1の統合手段で前記学習の結果を統合する際に用いられる重みを設定する予測重み付け手段を有する、請求項4に記載の能動学習システム。

【請求項7】

ラベル値が既知のデータを既知データとしラベル値が未知のデータを未知データとして前記既知データの集合と前記未知データの集合とを格納する記憶装置と、

既知データの学習及び未知データの予測を行う複数の学習機械と、

前記学習機械ごとに設けられ、前記記憶装置から前記既知データをサンプリングして対応する学習機械に入力する複数のサンプリング装置と、

前記各学習機械が前記既知データに基づいて行った学習の結果を統合する第1の統合手段と、

前記各学習機械が前記未知データに基づいて行った予測の結果から次に学習すべきデータを計算して出力する第2の統合手段と、

前記次に学習すべきデータに対応するラベル値を入力する結果入力手段と、

前記ラベル値が入力されたデータを前記未知データの集合から削除して前記既知データの集合に追加する制御手段と、

前記第2の統合手段で次に学習すべきデータを選択する際に用いられる重みを設定するデータ重み付け手段と、

を有する、能動学習システム。

【請求項8】

前記第2の統合手段で次に学習すべきデータを選択する際に用いられる重みを設定するデータ重み付け手段を有する、請求項4乃至6のいずれか1項に記載の能動学習システム。

【請求項9】

前記既知データ及び前記未知データにおいてグループ分けを行うグループ生成手段をさらに有し、

前記第2の統合手段は前記次に学習すべきデータを計算する際に、グループで散らばるようにデータを選択する、請求項4乃至8のいずれか1項に記載の能動学習システム。

【請求項10】

前記既知データ及び前記未知データにおいてグループ分けを行うグループ生成手段と、グループで散らばるように前記各学習機械に前記未知データを供給するデータ選択手段とをさらに有する、請求項4乃至8のいずれか1項に記載の能動学習システム。

【請求項11】

ラベル値が既知のデータを既知データとしラベル値が未知のデータを未知データとして前記既知データの集合と前記未知データの集合とを格納する記憶装置と、

既知データの学習及び未知データの予測を行う複数の学習機械と、

前記学習機械ごとに設けられ、前記記憶装置から前記既知データをサンプリングして対応する学習機械に入力する複数のサンプリング装置と、

前記各学習機械が前記既知データに基づいて行った学習の結果を統合する第1の統合手段と、

前記各学習機械が前記未知データに基づいて行った予測の結果から次に学習すべきデータを計算して出力する第2の統合手段と、

前記次に学習すべきデータに対応するラベル値を入力する結果入力手段と、

前記ラベル値が入力されたデータを前記未知データの集合から削除して前記既知データの集合に制御手段と、

前記既知データ及び前記未知データにおいてグループ分けを行うグループ生成手段と、  
を有し、前記第2の統合手段は前記次に学習すべきデータを計算する際に、グループで散らばるようにデータを選択する、能動学習システム。

【請求項12】

ラベル値が既知のデータを既知データとしラベル値が未知のデータを未知データとして前記既知データの集合と前記未知データの集合とを格納する記憶装置と、

既知データの学習及び未知データの予測を行う複数の学習機械と、

前記学習機械ごとに設けられ、前記記憶装置から前記既知データをサンプリングして対応する学習機械に入力する複数のサンプリング装置と、

前記各学習機械が前記既知データに基づいて行った学習の結果を統合する第1の統合手段と、

前記各学習機械が前記未知データに基づいて行った予測の結果から次に学習すべきデータを計算して出力する第2の統合手段と、

前記次に学習すべきデータに対応するラベル値を入力する結果入力手段と、

前記ラベル値が入力されたデータを前記未知データの集合から削除して前記既知データの集合に制御手段と、

前記既知データ及び前記未知データにおいてグループ分けを行うグループ生成手段と、

グループで散らばるように前記各学習機械に前記未知データを供給するデータ選択手段と、

を有する、能動学習システム。

【請求項13】

コンピュータを、

ラベル値が既知のデータを既知データとしラベル値が未知のデータを未知データとして前記既知データの集合と前記未知データの集合とを格納する記憶手段、

前記記憶手段から既知データをサンプリングするとともに既知データの学習及び未知データの予測を行う複数の学習手段、

前記各学習機械が前記既知データに基づいて行った学習の結果を統合する第1の統合手段、

前記各学習手段が前記未知データに基づいて行った予測の結果から次に学習すべきデータを計算して出力する第2の統合手段、

前記次に学習すべきデータに対応するラベル値を入力する結果入力手段、

前記ラベル値が入力されたデータを前記未知データの集合から削除して前記既知データの集合に追加する制御手段、

前記サンプリング装置でのサンプリング時の重み、前記第1の統合手段で用いられる重み、及び前記第2の統合手段で用いられる重みのうちの少なくとも1つの重みを設定する重み付け手段、

として機能させるプログラム。

【請求項14】

前記コンピュータをさらに、前記既知データ及び前記未知データにおいてグループ分けを行うグループ生成手段として機能させ、前記第2の統合手段において前記次に学習すべきデータを計算する際に、グループで散らばるようにデータが選択されるようにする、請求項13に記載のプログラム。

【請求項15】

前記コンピュータをさらに、

前記既知データ及び前記未知データにおいてグループ分けを行うグループ生成手段、

グループで散らばるように前記各学習機械に前記未知データを供給するデータ選択手段

、

として機能させる請求項13に記載のプログラム。

【請求項 16】

コンピュータが読み取り可能な記録媒体であって、請求項 13 乃至 15 のいずれか 1 項に記載のプログラムを格納した記録媒体。

【書類名】明細書

【発明の名称】能動学習方法およびシステム

【技術分野】

【0001】

本発明は、能動学習方法及び能動学習システムに関する。

【背景技術】

【0002】

能動学習とは、非特許文献1に記載されたところによれば、学習者が学習データを能動的に選択することによってできる学習形態であり、一般に学習を能動的に行うことにより、データ数や計算量の意味で学習の効率性を向上することができることが知られている。能動学習を行うシステムを能動学習システムと呼ぶ。例えば、集められたデータに対して統計的に分析を行い、ラベル値が未知のデータに対しては、過去のデータの傾向から結果の予測を行うような学習システムを考える。そのような学習システムには、能動学習システムを適用することができる。以下、この種の能動学習システムの概略を説明する。

【0003】

ラベル値が未知のデータとラベル値が既知のデータが存在するものとする。ラベル値が既知のデータで学習を行い、その学習した結果をラベル値が未知のデータに適用する。そのときに、ラベル値が未知のデータから効率的に学習を行えるようなデータを学習システムが選択し、そのデータを出力する。出力されたデータに対して実験するなり調査するなりして、ラベル値が未知のデータに対する結果を得て、それを入力したのちラベル値が既知のデータに混ぜて、同じように学習を行う。その一方で、ラベル値が未知のデータの集合からは、結果が得られたデータを削除する、というものである。能動学習システムでは、そのような動作を繰り返して行っていく。

【0004】

また、データは以下のように記述されている。1つのデータは、複数の属性とラベルというもので記述される。たとえば有名な評価データの中には、“golf”というものがある。それはゴルフをプレーするかしないかを判定するものであって、天気、温度、湿度、風の強さという4つのものから記述されている。天気は、「晴れ」、「曇り」または「雨」、風は、「有」または「無」という値をとる。気温と湿度は実数値である。たとえば1つのデータは、天気：晴れ、温度：15度、湿度：40%、風：無、プレー：する、というように書かれている。そのデータの場合、天気、温度、湿度、風の4つを属性と呼ぶ。また、プレーする、しないという結果のことをラベルと呼ぶ。本明細書では、ラベルがとりうる値が離散値の場合には、特にクラスと呼ぶ。

【0005】

ここで、さまざまな用語を定義しておく。

【0006】

仮にラベルは2値であるとしておく。その2値のうち注目しているラベルの方を正例、それ以外のものを負例とする。またラベルが多値の場合には、注目している1つのラベル値を正例、それ以外のすべてのラベル値を負例とする。またラベルがとりうる値が連続値の場合には注目する値付近にラベル値が存在するとき正例と呼び、それ以外のところにあるときに負例と呼ぶことにする。

【0007】

学習の精度を測る指標としては、ROC（受信者動作特性：receiver operating characteristic）曲線、ヒット率、正解率の推移などがある。以下の説明では、これら3つの指標を用いて評価を行う。

【0008】

ROC曲線は、以下のように定義される。

横軸：負例のうち正例と判断されたデータの個数／全負例数、

縦軸：正例のうち正例と判断されたデータの個数／全正例数。

ランダムな予測を行ったとき、ROC曲線は、原点と（1，1）を結ぶ対角線となる。

## 【0009】

ヒット率は以下のように定義される。

横軸：ラベル値が既知のデータ数 / (ラベル値が未知 + 既知のデータ数) ,

縦軸：ラベル値が既知のデータの中の正例数 / 全正例数。

ランダムな予測を行ったとき。ヒット率は、原点と (1, 1) を結ぶ対角線となる。また、限界は原点と (正例数 / (ラベル値が未知 + 既知のデータ数), 1) を結んだ線となる。

## 【0010】

正解率の推移は以下のように定義される。

横軸：ラベル値が既知のデータ数,

縦軸：正しく判断されたデータの個数 / ラベル値が既知のデータ数。

後述する「発明を実施するための最良の形態」においては、これらの指標を用いて、本発明による能動学習システムを評価している (図 3、5、7、9、11、13、15、18 を参照)。

## 【0011】

またエントロピーとは、以下のように定義される。各  $P_i$  は  $i$  である確率を示しているとする。

## 【0012】

エントロピー  $= -(p_1 * \log(P_1) + p_2 * \log(P_2) + \dots + P_n * \log(P_n))$

なお、従来の能動学習システムを開示するものとしては、特開平 11-316754 号公報 (特許文献 1) に開示されたものがある。この公報に開示の能動学習システムは、学習の精度を向上するために、下位アルゴリズムに学習を行わせる学習段階と、学習精度をブースティングにより向上させるブースティング段階と、複数の入力候補点に対する関数値予測段階と、重みの総和が最大の出力値の重み和と、重みの総和が次に大きい出力値の重み和との差が最も小さいような入力点を選択する入力点指定段階と、を行うことを特徴とするものである。

## 【0013】

また、非特許文献 1 (「発見科学とデータマイニング」, 共立出版、2001 年 6 月、pp. 64-71) には、複数の学習機械を備えるシステムを使用し、各学習機械はデータからランダムにサンプリングしてそのデータを学習し、ラベル値が未知のデータに対してはそれぞれの学習機械が予測を行ってもっとも分散が大きくなるような点を次に学習すべき点として出力するような手法を開示している。

【特許文献 1】特開平 11-316754 号公報

【非特許文献 1】「発見科学とデータマイニング」, 森下真一・宮野 悟編、共立出版、2001 年 6 月、ISBN4-320-12018-3、pp. 64-71

## 【発明の開示】

## 【発明が解決しようとする課題】

## 【0014】

しかしながら上述した従来の手法は、次に学習すべきデータとして出力するものをユーザーの意思で制御できない、という問題点を有する。その理由は、これらの従来の手法は、なるべく早く学習精度を向上することを目標とするために、次の候補点として最も分散の大きな点や、下位の学習機械からの出力が割れる点のみを選んでからである。

## 【0015】

これらの従来の手法は、学習データにおいて、対象となる値やクラスのデータ数が他のクラスや値のものに比べてきわめて低い状況下において、対象となっている値やクラスの正解率を得るのが困難である、という問題点を有する。その理由は、今までに開発されてきた下位学習アルゴリズムが極端な個数の不平等な状況までを考察して設計されたものではないことにあるばかりでなく、従来の能動学習アルゴリズムも同じようにそのような状況を想定していないことにある。



## 【0016】

またこれらの従来の手法が用いる能動学習アルゴリズムは、入力すべきデータの選択の段階において、似たようなデータを入力点として数多く出力してしまう、という問題点がある。その理由もやはり、従来の能動学習アルゴリズムが、下位の学習アルゴリズムが学習したものを十分に活かせるようなメカニズムを有していないことにある。

## 【0017】

さらにこれらの従来の手法は、システムにおいて最終判断の方法が予め定められているために、学習の精度を変えられない、という問題点も有する。

## 【0018】

そこで本発明の目的は、能動学習法の精度を向上させつつ、利用者の意思で精度を制御することができ、また、興味のあるデータを先に抜き出すという機能なども備える能動学習方法及びシステムを提供することにある。

## 【課題を解決するための手段】

## 【0019】

本発明の能動学習方法は、ラベル値が既知のデータを既知データとしラベル値が未知のデータを未知データとして既知データの集合と未知データの集合とを格納する記憶装置と、複数の学習機械とを使用する能動学習方法であって、複数の学習機械が、既知データに関し、記憶装置からそれぞれ独立にサンプリングを行った後に学習を行う段階と、学習の結果として、複数の学習機械の出力結果を統合して出力する段階と、複数の学習機械が、記憶装置から未知データを取り出して予測を行う段階と、予測の結果に基づいて次に学習すべきデータを計算して出力する段階と、次に学習すべきデータに対応するラベル値を入力する段階と、ラベル値が入力されたデータを未知データの集合から削除して既知データの集合に追加する段階と、を有し、既知データをサンプリングするとき、複数の学習機械による学習の結果を統合するとき、及び、複数の学習機械による予測から次に学習すべきデータを計算するとき、のうちの少なくとも1つにおいて、均等でない重み付けを実行する。

## 【0020】

本発明の能動学習方法における重み付けでは、例えば、データ数に極端に偏りが発生しているときに、その比重を重くするようにする。さらにこの方法では、次に予測のために学習装置に入力すべきデータの選択において、選ばれてきた候補のデータの中から空間的なデータの分布を考慮に入れながらさらに選び出すというデータの分布に広がりを持たせる機構を付け加えることによって、お互いに似たようなデータを出力することを避けることができる。

## 【0021】

本発明の能動学習システムは、ラベル値が既知のデータを既知データとしラベル値が未知のデータを未知データとして既知データの集合と未知データの集合とを格納する記憶装置と、既知データの学習及び未知データの予測を行う複数の学習機械と、学習機械ごとに設けられ、記憶装置から既知データをサンプリングして対応する学習機械に入力する複数のサンプリング装置と、各学習機械が既知データに基づいて行った学習の結果を統合する第1の統合手段と、各学習機械が未知データに基づいて行った予測の結果から次に学習すべきデータを計算して出力する第2の統合手段と、次に学習すべきデータに対応するラベル値を入力する結果入力手段と、ラベル値が入力されたデータを未知データの集合から削除して既知データの集合に追加する制御手段と、を有するとともに、(1) サンプリング装置ごとにサンプリング時の重みを設定するサンプリング重み付け手段、(2) 第1の統合手段で学習の結果を統合する際に用いられる重みを設定する予測重み付け手段、(3) 第2の統合手段で次に学習すべきデータを選択する際に用いられる重みを設定するデータ重み付け手段、及び(4) 既知データ及び未知データにおいてグループ分けを行うグループ生成手段のうちの少なくとも1つを有する。

## 【0022】

本発明の能動学習システムにおける重み付けでは、例えば、データ数に極端に偏りが発

生しているときに、その比重を重くするようにする。

【発明の効果】

【0023】

本発明によれば、(1) 学習データをサンプリングする際にデータに重み付けを行う、(2) 入力候補点から入力点を選び出す際にデータに重み付けを行う、及び(3) 入力されたデータに対して予測を行う際にデータに重み付けを行う、の計3通りの重み付けのうちの少なくとも1つを採用することにより、以下のような効果が得られる。

【0024】

まず第1に、対象となっているデータの重みを重くすることで重点的に学習を行うことができるので、対象となっているデータの、全体に対する割合が極めて低い状況下において、マイニングの精度を向上させながら、カバー率を向上させることができる。

【0025】

第2に、対象となっているデータの重みを軽くすることで、対象となっているものが未発見の領域の学習を行っていくことが可能になるため、対象となっているデータがさまざまな特徴に基づいているとき、それを、早期に発見できる。

【0026】

第3に、従来の能動学習法がデータの予測を行うときに均等に学習結果を扱っているのに対し、本発明では重み付けすることが可能なため、重みを変えることで精度を制御することが可能になり、任意の精度で学習することが可能になる。

【0027】

さらに、従来の方法では次に学習すべきデータが空間的にある領域に固まる傾向にあるが、本発明では、それらのデータが空間的に散らばるような機構を設けることにより、従来の能動学習法の欠点を修正でき、正解率を従来のものよりも高めることができる。

【発明を実施するための最良の形態】

【0028】

次に、本発明の好ましい実施の形態について、図面を参照して詳細に説明する。本発明の能動学習システムは、(1) 学習データをサンプリングする際にデータに重み付けを行う、(2) 入力候補点から入力点を選び出す際にデータに重み付けを行う、及び(3) 入力されたデータに対して予測を行う際にデータに重み付けを行う、の計3通りの重み付けのうちの少なくとも1つを採用することにより、上述した本発明の目的を達しようとするものである。これらの重み付けにおいては、データ数に極端に偏りが発生しているときに、その比重を重くするように重み付けを行っている。本発明は、どの段階で重み付けを行うかによって種々の実施形態が考えられる。

【0029】

また、本発明では、次に学習すべきデータを選択する際に、選ばれてきた候補のデータから空間的なデータの分布を考慮に入れながらさらに選び出すという、データの分布に広がりを持たせるメカニズムを付け加えることによって、お互いに似たようなデータを出力することを避けることができる。このようなメカニズムの有無によっても、本発明は種々の実施形態が考えられる。

【0030】

以下、このような各種の実施形態について説明する。

【0031】

《第1の実施形態》

図1は、本発明の第1の実施形態の能動学習システムの構成を示している。この能動学習システムは、ラベル値が既にわかっているデータ（すなわち既知データ）を蓄えておく記憶装置101と、記憶装置101内の既知データをサンプリングする時に重み付けを行うためのデータを生成するサンプリング重み付け装置102と、予測を行う時に重み付けを行うためのデータを生成する予測重み付け装置103と、次に学習すべきデータを選択するときに重み付けを行うためのデータを生成するデータ重み付け装置104と、複数の学習機械106と、記憶装置101からデータをサンプリングして対応する学習機械10

6 に対してデータを供給する複数のサンプリング装置 105 と、複数の学習機械 106 からの学習結果をまとめる規則統合装置 107 と、規則統合装置 107 に接続した出力装置 111 と、複数の学習機械 106 での結果に基づき次に学習すべきデータを計算するデータ統合装置 108 と、データ統合装置 108 に接続した出力装置 112 と、次に学習すべきデータの出力結果に対して結果を入力する結果入力装置 113 と、ラベル値がわかっていないデータ（すなわち未知データ）を格納する記憶装置 109 と、この能動学習システム全体の制御を行う制御装置 110 と、を備えている。特に制御装置 110 は、結果入力装置 113 によって入力された結果を例えば表形式のものとしてまとめ、記憶装置 109 内における該当するデータを削除し、その代わりに、結果が加えられたデータを記憶装置 101 内に格納するという制御を行う。サンプリング装置 105 と学習機械 106 とは、ここでは、1 対 1 の関係で設けられている。また、各学習機械 106 には、対応するサンプリング装置 105 からデータが供給されるとともに、記憶装置 109 からは未知データが供給される。規則統合装置 107 に接続した出力装置 111 からは、学習した規則が出力され、データ統合装置 108 に接続した出力装置 112 からは、次に学習すべきデータが出力される。

#### 【0032】

サンプリング重み付け装置 102 は、各サンプリング装置 105 に対して、記憶装置 101 に格納された既知データに基づき、サンプリング時にデータに対して重み付けを行うための重み付けデータを生成して供給する。予測重み付け装置 103 は、記憶装置 101 に格納された既知データに基づき、規則統合装置 107 において各学習機械 106 ごとの学習結果をまとめる際に重み付けを行うための重み付けデータを生成し、生成した重み付けデータを規則統合装置 107 に供給する。同様にデータ重み付け装置 104 は、記憶装置 101 に格納された既知データに基づき、データ統合装置 108 において次に学習すべきデータを選択して出力する際に重み付けを行うための重み付けデータを生成し、生成した重み付けデータをデータ統合装置 108 に供給する。

#### 【0033】

ここで、サンプリング重み付け装置 102、予測重み付け装置 103 及びデータ重み付け装置 104 のそれぞれによる重み付けを説明する。これらの重み付け装置 102～104 による重み付けとしては、均等でない重み付けであれば種々のものを使用できる。

#### 【0034】

サンプリング重み付け装置 102 での重み付けとしては、例えば、（１）既知データにおけるクラスあるいは値に応じた重みを設定する、（２）ラベル値が離散値を取る場合に、各サンプリング装置 105 が、独立して、あるクラスのデータの全てとそれ以外のクラスのデータからランダムにサンプリングするように重みを設定する、（３）ラベル値が連続値を取る場合に、各サンプリング装置 105 が、特定のラベルの値の付近のデータの全てとそれ以外のラベル値のデータからランダムにサンプリングするように重みを設定する、などが挙げられる。

#### 【0035】

予測重み付け装置 103 での重み付けとしては、例えば、各学習機械 106 が出力した結果についてクラス（ラベル値が離散値を取る場合）ごとあるいは数値における区間（ラベル値が連続値を取る場合）ごとに重みを決定する方法がある。

#### 【0036】

データ重み付け装置 104 での重み付けとしては、例えば、（１）ラベル値が離散値をとる場合にクラスごとの頻度から計算されるばらつき具合に応じて重みを割当てる、（２）各学習機械 106 での結果として得られる値の分散に応じて重みを割当てる、（３）ラベル値が離散値をとる場合にクラスごとの頻度から計算されるエントロピーに応じて重みを割当てる、などが挙げられる。ばらつき具合に応じて重みを割当てる場合には、ばらつき具合が最大のところのみが最大の重みになることは除くようにしてもよい。同様に、分散あるいはエントロピーに応じて重みを割当てる場合には、それら分散あるいはエントロピーが最大のところのみが最大の重みになることは除くようにしてもよい。さらに、これ

らの重みとは別個に各学習機械 106 で得られる結果そのものに対して重みを割当てるようにしてもよい。

#### 【0037】

次に、本実施形態の能動学習システムの動作について、図 2 に示すフローチャートを利用して説明する。ここでは、データは表形式で与えられるものとする。

#### 【0038】

まず、ステップ 201 において、ラベル値が既知のデータは記憶装置 101 に、ラベル値が未知のデータは記憶装置 109 に記憶される。その結果、既知データの集合が記憶装置 101 に格納され、未知データの集合が記憶装置 109 に格納されることになる。

#### 【0039】

次に、ステップ 202 において、サンプリング重み付け装置 102 は、記憶装置 101 から送られてきたデータに基づいて重み（すなわち重み付けデータ）を生成し、あるいはそのような重みを読み込み、各サンプリング装置 105 に送る。各サンプリング装置 105 は、サンプリング重み付け装置 102 から送られてきた重みにしたがって重み付けを行いながら、記憶装置 101 内の既知データをサンプリングし、サンプリングしたデータを対応する学習機械 106 に送る。各学習機械 106 は、ステップ 203 において、サンプリング装置から送られてきたデータに基づいて学習を実行する。

#### 【0040】

記憶装置 101 からは予測重み付け装置 103 にもデータが送られており、ステップ 204 において、予測重み付け装置 103 は、記憶装置 101 から送られてきたデータに基づいて重み（すなわち重み付けデータ）を生成し、あるいはそのような重みを読み込み、それらを規則統合装置 107 へ送る。規則統合装置 107 は、重み付けデータに基づいて、各学習機械 106 からの学習結果に重み付けを行いながらこれらの学習結果をまとめる。このとき、各学習機械 106 が出力した結果についてクラス（ラベル値が離散値を取る場合）ごとあるいは数値における区間（ラベル値が連続値を取る場合）ごとに頻度を計算し、頻度と上述した重みとを乗算し、その値がもっとも大きな値となっているものを予想値として出力する。規則統合装置 107 は、学習結果をまとめた結果を規則として出力装置 111 に送る。

#### 【0041】

次に各学習機械 106 は、ステップ 205 において、記憶装置 109 に格納されたラベル値が未知のデータに対して予測を行い、その結果は、データ統合装置 108 に送られる。このとき、記憶装置 101 からはデータ重み付け装置 104 にもデータが送られており、ステップ 206 において、データ重み付け装置 104 は、記憶装置 101 から送られてきたデータに基づいて重み（すなわち重み付けデータ）を生成し、あるいはそのような重みを読み込み、それらをデータ統合装置 108 へ送る。データ統合装置 108 は、重み付けデータに基づいて、各学習機械 106 からの予測結果に重み付けを行いながらこれらの結果をまとめ、次に学習すべきデータを選択する。次に学習すべきデータを選択方法としては、以下のようなものが挙げられる。例えば、（1）ばらつき具合あるいはエントロピーに応じて重みが割当てられている場合には、各学習機械 106 が出力した結果からそのクラスごとに頻度を計算し、頻度を元にばらつき具合あるいはエントロピーを示す数値を計算し、ばらつき具合あるいはエントロピーに応じて割り当てられた重みが重い順番にデータを選択する、（2）分散に応じて重みが割当てられている場合には、各学習機械 106 が出力した結果からその分散を計算し、分散に応じて割り当てられた重みが重い順番にデータを選択する、（3）ばらつき具合あるいはエントロピーと結果とのそれぞれに応じて重みが割当てられている場合には、各学習機械 106 が出力した結果からそのクラスごとに頻度を計算し、頻度を元にばらつき具合あるいはエントロピーを示す数値を計算し、ばらつき具合あるいはエントロピーに応じて割り当てられた重みと結果に割り当てられた重みとをあわせて重みの重い順番にデータを選択する、（4）分散と結果のそれぞれに応じて重みが割当てられている場合には、各学習機械 106 が出力した結果からその分散を計算し、分散に応じて割り当てられた重みと結果に割り当てられた重みとをあわせて重み

の重い順番にデータを選択する。データ統合装置108は、その結果を次に学習すべきデータとして出力装置112に送る。

#### 【0042】

次に、ステップ207において、次に学習すべきデータに対する結果（ラベル値）が、結果入力装置113を介し、人手によって、あるいはコンピュータにより入力される。入力された結果は制御装置110に送られ、制御装置110は、その結果が入力されたデータを記憶装置109から削除し、その代わりに記憶装置101に記憶させる。

#### 【0043】

以後、上述の処理が繰り返され、能動学習が進行する。この場合、これらの処理は、最長の場合で記憶装置109内に未知データがなくなるまで行われるが、その前に打ち切るようにしてもよい。後述するように本実施形態によれば、迅速に“よい結果”を得ることができるので、例えば適当な反復回数を設定してそこで処理を打ち切るようにすることができる。サンプリング重み付け装置102、予測重み付け装置103及びデータ重み付け装置104は、いずれも均等でない重み付けを行う。

#### 【0044】

図3(a)～(c)は、本実施形態の能動学習システムの効果を説明する図である。図3(a)において破線301は従来の能動学習法を用いた場合のヒット率を示し、実線302は、本実施形態の能動学習システムを用いた場合のヒット率を示している。本実施形態によれば、従来の能動学習法よりも早期の段階において、対象となっているクラス（値）のデータを見つけていることがわかる。図3(b)において、破線303は従来の能動学習法を用いた場合のROC曲線を示し、実線304は本実施形態の能動学習システムを用いた場合のROC曲線を示している。本実施形態によれば、従来の能動学習法に比べて高い精度で学習を行えていることがわかる。さらに、従来の能動学習法ではどのような精度にするのかを外部から制御することができなかったが（曲線上のある1点に存在する）、本実施形態の手法では、予測重み付け装置103における重みを変えることで、任意の精度（図では、A、B、C、Dと4つの直線が存在するが、任意の位置に対する例）を設定できる。さらに図3(c)において、破線305は従来の能動学習法を用いた場合の正解率の推移を示し、破線306は、本実施形態の能動学習システムを用いた場合における正解率の推移を示している。本実施形態によれば、対象となっているクラス（値）のデータの重みを重くすることによって、そのクラスに関する正解率を上げることができることがわかる。

#### 【0045】

##### 《第2の実施形態》

次に、本発明の第2の実施形態について、図4を参照して説明する。図4に示す能動学習システムは、第1の実施形態の能動学習システムと同様のものであるが、予測重み付け装置とデータ重み付け装置とが設けられていない点で、第1の実施形態のものと相違する。予測重み付け装置とデータ重み付け装置とが設けられていないことにより、規則統合装置107では、学習機械106から出てきた結果がすべて均等に取り扱われ、多数決などの手段によって最終的な規則が出力されることになる。具体的には、規則統合装置107は、各学習機械106が出力した結果についてクラス（ラベル値が離散値を取る場合）ごとあるいは数値における区間（ラベル値が連続値を取る場合）ごとに頻度を計算し、その値がもっとも大きな値となっているものを予想値として出力する。

#### 【0046】

またデータ統合装置108においも同様に出力結果が均等に扱われ、もっとも判断に迷うデータが出力されることになる。具体的には、例えば、(1)ラベル値が離散値を取る場合に、各学習機械106が出力した結果からそのクラスごとに頻度を計算し、頻度を元にばらつき具合を示す数値を計算し、あるクラスと判断されたデータとばらつき具合を示す指標が最大もしくは最大付近のデータから、次に学習すべきデータを選択する。(2)ラベル値が連続値を取る場合に、各学習機械106が出力した結果からその分散を計算し、ある数値の付近にあるデータと分散が最大もしくは最大付近のデータから、次に学習す

べきデータを選択する。(3) 各学習機械 106 が出力した結果からその分散を計算し、特定のクラス以外のデータ (あるいはある数値付近にないデータ) であってかつ “分散が最小もしくは最小に近い” データから、次に学習すべきデータを選択する。

#### 【0047】

図 5 は、本実施形態の能動学習システムの効果を示す図である。図において、破線 307 は、従来の能動学習法を用いた場合の学習精度を表す ROC 曲線であり、実線 308 は、本実施形態の能動学習システムにより、対象となっているクラス (値) のデータが多く選ばれるようなサンプリングをしたとき学習精度を示す ROC 曲線である。本実施形態によれば、従来の能動学習法よりも高い精度が得られることがわかる。

#### 【0048】

##### 《第 3 の実施形態》

次に、本発明の第 3 の実施形態について、図 6 を参照して説明する。図 6 に示す能動学習システムは、第 1 の実施形態の能動学習システムと同様のものであるが、サンプリング重み付け装置とデータ重み付け装置とが設けられていない点で、第 1 の実施形態のものと相違する。サンプリング重み付け装置とデータ重み付け装置とが設けられていないことにより、各サンプリング装置 105 では既知のデータがすべて均等に取り扱われ、ランダムなサンプリングが行われる。またデータ統合装置 108 では、第 2 の実施形態の場合と同様に、出力結果が均等に扱われ、もっとも判断に迷うデータが出力されることになる。

#### 【0049】

図 7 は、本実施形態の能動学習システムの効果を示す図であり、図において線 309 は、この能動学習システムの学習精度を表す ROC 曲線を示している。従来の能動学習法では学習結果を統合するときに均等に結果を扱っていたため、ある特定の精度でしか能動学習システムを構築することができなかった。本実施形態によれば、任意の重みで学習結果を統合することができるため、たとえば A、B、C、D のような精度でシステムを構成することができる。

#### 【0050】

##### 《第 4 の実施形態》

次に、本発明の第 4 の実施形態について、図 8 を参照して説明する。図 8 に示す能動学習システムは、第 1 の実施形態の能動学習システムと同様のものであるが、サンプリング重み付け装置と予測重み付け装置とが設けられていない点で、第 1 の実施形態のものと相違する。サンプリング重み付け装置と予測重み付け装置とが設けられていないことにより、各サンプリング装置 105 では既知のデータがすべて均等に取り扱われ、ランダムなサンプリングが行われる。また、規則統合装置 107 では、第 2 の実施形態の場合と同様に、学習機械 106 から出てきた結果がすべて均等に取り扱われ、多数決などの手段によって最終的な規則が出力されることになる。

#### 【0051】

図 9 は、本実施形態の能動学習システムの効果を示す図である。図において、破線 310 は従来の能動学習法を用いた場合の正解率の推移を示し、破線 311 は、本実施形態の能動学習システムを用いた場合における正解率の推移を示している。本実施形態では、サンプリングのときの重みとして、次に実験すべきデータがなるべく散らばるような重み付けを行っている。このような重み付けを用いることにより、従来の能動学習法よりも早く学習していることわかる。

#### 【0052】

##### 《第 5 の実施形態》

次に、本発明の第 5 の実施形態について、図 10 を参照して説明する。図 10 に示す能動学習システムは、第 1 の実施形態の能動学習システムと同様のものであるが、データ重み付け装置が設けられていない点で、第 1 の実施形態のものと相違する。データ重み付け装置が設けられていないことにより、データ統合装置 108 では、第 2 の実施形態の場合と同様に、出力結果が均等に扱われ、もっとも判断に迷うデータが出力されることになる。

。



## 【0053】

図11は、本実施形態の能動学習システムの効果を示す図である。図において、破線312は従来の能動学習法を用いた場合のROC曲線を示し、破線313は、本実施形態の能動学習システムを用いた場合におけるROC曲線を示している。本実施形態では、サンプリング時には、あるクラス（値）の重みが重くなるような重み付けを行い、次に学習すべきデータを選択する時には、同様に、そのクラスの重みが重くなるような重み付けを行っている。図11から分かるように、本実施形態によれば、学習の精度が向上し、また、予測重み付け装置の重みを変えることで、A、B、C、Dにあるように、任意の精度で学習を行うことができるようになる。

## 【0054】

## 《第6の実施形態》

次に、本発明の第6の実施形態について、図12を参照して説明する。図12に示す能動学習システムは、第1の実施形態の能動学習システムと同様のものであるが、予測重み付け装置が設けられていない点で、第1の実施形態のものと相違する。予測重み付け装置が設けられていないことにより、規則統合装置107では、第2の実施形態の場合と同様に、学習機械106から出てきた結果がすべて均等に取り扱われ、多数決などの手段によって最終的な規則が出力されることになる。

## 【0055】

図13(a)、(b)は、本実施形態の能動学習システムの効果を示す図である。図13(a)において、破線314は従来の能動学習法を用いた場合のヒット率を示し、実線315は、本実施形態の能動学習システムを用いた場合におけるヒット率を示している。また、図13(b)において、破線316は従来の能動学習法を用いた場合のROC曲線を示し、実線317は、本実施形態の能動学習システムを用いた場合のROC曲線を示している。本実施形態においては、サンプリング時には、あるクラス（値）の重みが重くなるように重み付けを行っており、次に学習すべきデータを選択するときも、同様に、そのクラスの重みが重くなるように重み付けを行っている。本実施形態によれば、対象となるクラス（値）の9割を従来のものよりも早く発見でき、また、学習精度も向上していることがわかる。

## 【0056】

## 《第7の実施形態》

次に、本発明の第7の実施形態について、図14を参照して説明する。図14に示す能動学習システムは、第1の実施形態の能動学習システムと同様のものであるが、サンプリング重み付け装置が設けられていない点で、第1の実施形態のものと相違する。サンプリング重み付け装置が設けられていないことにより、各サンプリング装置105では既知のデータがすべて均等に取り扱われ、ランダムなサンプリングが行われる。

## 【0057】

図15(a)、(b)は、本実施形態の能動学習システムの効果を示す図である。図15(a)において、破線318は従来の能動学習法を用いた場合のヒット率を示す、実線19は、本実施形態の能動学習システムを用いた場合のヒット率を示している。図15(b)において、破線320は、本実施形態の能動学習システムを用いた場合におけるROC曲線を示している。本実施形態では、次に学習すべきデータを選択するときの重み付けも、学習結果を統合するときの重み付けも、あるクラス（値）のデータの重みが重くなるようにした。本実施形態によれば、重みを重くしたクラスのデータが早く出力されており、また図示A、B、C、Dに示されるように、任意の精度で学習を行えるようになっている。

## 【0058】

## 《第8の実施形態》

次に、本発明の第8の実施形態について、図16を参照して説明する。図16に示す能動学習システムは、第1の実施形態の能動学習システムと同様のものであるが、グループ生成装置115が付加されているとともに、データ統合装置とそのデータ統合装置に接続

する出力装置とがデータ統合選択装置 114 で置き換えられている点で相違する。データ統合選択装置 114 は、第 1 の実施形態のシステム（図 1 参照）におけるデータ統合装置 108 と出力装置 112 の機能を合わせ持つものであるが、次に学習すべきデータを選択する際に、グループ生成装置 115 でのグループ分けにしたがって、相互のデータがなるべくグループに散らばるように選択する。グループ選択装置 114 は、記憶装置 101 に格納されたラベル値が既知のデータ、または記憶装置 109 に格納されたラベル値が未知のデータ、もしくはその両者のデータをグループ分けするものである。

#### 【0059】

次に、本実施形態の能動学習システムの動作について、図 17 に示すフローチャートを利用して説明する。ここでは、データは表形式で与えられるものとする。

#### 【0060】

まず、ステップ 211 において、ラベル値が既知のデータは記憶装置 101 に、ラベル値が未知のデータは記憶装置 109 に記憶される。グループ生成装置 115 は、ステップ 212 において、記憶装置 101 内の既知データと記憶装置 109 内の未知データに関してグループ分けを行う。グループ分けの結果は、グループ生成装置 115 からグループ情報として出力される。

#### 【0061】

次に、ステップ 213 において、サンプリング重み付け装置 102 は、記憶装置 101 から送られてきたデータに基づいて重み（すなわち重み付けデータ）を生成し、あるいはそのような重みを読み込み、各サンプリング装置 105 に送る。各サンプリング装置 105 は、サンプリング重み付け装置 102 から送られてきた重みにしたがって重み付けを行いながら、記憶装置 101 内の既知データをサンプリングし、サンプリングしたデータに対応する学習機械 106 に送る。各学習機械 106 は、ステップ 214 において、サンプリング装置から送られてきたデータに基づいて学習を実行する。

#### 【0062】

記憶装置 101 からは予測重み付け装置 103 にもデータが送られており、ステップ 215 において、予測重み付け装置 103 は、記憶装置 101 から送られてきたデータに基づいて重み（すなわち重み付けデータ）を生成し、あるいはそのような重みを読み込み、それらを規則統合装置 107 へ送る。規則統合装置 107 は、重み付けデータに基づいて、各学習機械 106 からの学習結果に重み付けを行いながらこれらの学習結果をまとめる。規則統合装置 107 は、学習結果をまとめた結果を規則として出力装置 111 に送る。

#### 【0063】

次に各学習機械 106 は、ステップ 216 において、記憶装置 109 に格納されたラベル値が未知のデータに対して予測を行い、その結果は、データ統合選択装置 114 に送られる。

#### 【0064】

このとき、記憶装置 101 からはデータ重み付け装置 104 にもデータが送られており、ステップ 217 において、データ重み付け装置 104 は、記憶装置 101 から送られてきたデータに基づいて重み（すなわち重み付けデータ）を生成し、あるいはそのような重みを読み込み、それらをデータ統合選択装置 114 へ送る。データ統合選択装置 114 は、重み付けデータとグループ生成装置 115 からのグループ情報とに基づき、各学習機械 106 からの予測結果に重み付けを行いながらこれらの結果をまとめ、次に学習すべきデータを選択する。その際、データ統合選択装置 114 は、グループ生成装置 814 でのグループ分けにしたがって、お互いのデータがなるべくグループに散らばるように、次に学習すべきデータをする。

#### 【0065】

次に、ステップ 218 において、次に学習すべきデータに対する結果（ラベル値）が、結果入力装置 113 を介し、人手によって、あるいはコンピュータにより入力される。入力された結果は制御装置 110 に送られ、制御装置 110 は、その結果が入力されたデータを記憶装置 109 から削除し、その代わりに記憶装置 101 に記憶させる。以後、第 1



の実施形態の場合と同様に、上述の処理が繰り返され、能動学習が進行する。

#### 【0066】

図18は、本実施形態の能動学習システムの効果を説明する図である。図において、破線321は従来の能動学習法を用いた場合の正解率の推移を示し、破線322は、第1の実施形態の能動学習システムを用いた場合における正解率の推移を示し、実線323は、グループ生成装置115によって作成されたグループ情報をもとにして次に学習すべきデータを選択する本実施形態の能動学習システムを用いた場合における正解率の推移を示している。グループ生成装置によって生成されたグループの情報をもとにして、次に学習すべきデータを選択する際にお互いのデータがなるべく異なったグループに属するようにデータを選択することで、正解率を早い段階で高くすることができることがわかる。

#### 【0067】

なお、本実施形態は、サンプリング重み付け装置102、予測重み付け装置103及びデータ重み付け装置104のうちの一部または全部を設けない構成とすることもできる。

#### 【0068】

##### 《第9の実施形態》

次に、本発明の第9の実施形態について、図19を参照して説明する。図19に示す能動学習システムは、第8の実施形態の能動学習システムと同様のものであるが、データ選択装置116が新たに設けられ、また、データ統合選択装置の代わりに第1の実施形態の場合と同様のデータ統合装置108及び出力装置112が設けられている点で、第8の実施形態のものと相違する。データ選択装置118は、グループ生成装置115からのグループ情報にしたがって、各学習機械106での予測の対象となる未知データを記憶装置109から選択し、選択された未知データを各学習機械106に送るものである。

#### 【0069】

この能動学習システムでは、グループ生成装置115で生成されたグループは、データ選択装置116に送られる。記憶装置109からは未知データがデータ選択装置116に送られる。データ選択装置116は、なるべく異なったグループに散らばるように未知データが選択して、選択されたデータが、予測のために学習機械106に送られる。データ統合装置108は、データ重み付け装置904で決定された重み付けを適用して、次に学習すべきデータを選択する。この能動学習システムは、第8の実施形態の能動学習システムと同様の効果を奏する。

#### 【0070】

なお、本実施形態は、サンプリング重み付け装置102、予測重み付け装置103及びデータ重み付け装置104のうちの一部または全部を設けない構成とすることもできる。

#### 【0071】

以上説明した能動学習システムは、それを実現するためのコンピュータプログラムを、パーソナルコンピュータやワークステーションなどのコンピュータに読み込ませ、そのプログラムを実行させることによって実現できる。能動学習を行うためのプログラム（能動学習システム用プログラム）は、磁気テープやCD-ROMなどの記録媒体によって、あるいはネットワークを介して、コンピュータに読み込まれる。そのようなコンピュータは、一般に、CPUと、プログラムやデータを格納するためのハードディスク装置と、主メモリと、キーボードやマウスなどの入力装置と、CRTや液晶ディスプレイなどの表示装置と、磁気テープやCD-ROM等の記録媒体を読み取る読み取り装置と、ネットワークとのインタフェースとなる通信インタフェースとから構成されている。ハードディスク装置、主メモリ、入力装置、表示装置、読み取り装置及び通信インタフェースは、いずれもCPUに接続している。このコンピュータでは、能動学習を実行するためのプログラムを格納した記録媒体を読み取り装置に装着し、記録媒体からプログラムを読み出してハードディスク装置に格納し、あるいはそのようなプログラムをネットワークからダウンロードしてハードディスク装置に格納し、その後、ハードディスク装置に格納されたプログラムをCPUが実行することにより、上述した能動学習が実行されることになる。

#### 【図面の簡単な説明】

## 【0072】

【図1】本発明の第1の実施形態の能動学習システムの構成を示すブロック図である。

【図2】図1に示すシステムを用いた能動学習法の処理を示すフローチャートである。

【図3】図1に示すシステムによる能動学習法と従来の能動学習法とにおける学習精度を比較するグラフであって、(a)はヒット率を示すグラフ、(b)はROC曲線を示すグラフ、(c)は正解率の推移を示すグラフである。

【図4】本発明の第2の実施形態の能動学習システムの構成を示すブロック図である。

【図5】図4に示すシステムによる能動学習法と従来の能動学習法とにおける学習精度を比較するROC曲線のグラフである。

【図6】本発明の第3の実施形態の能動学習システムの構成を示すブロック図である。

【図7】図6に示すシステムによる能動学習法と従来の能動学習法とにおける学習精度を比較するROC曲線のグラフである。

【図8】本発明の第4の実施形態の能動学習システムの構成を示すブロック図である。

【図9】図8に示すシステムによる能動学習法と従来の能動学習法とにおける学習精度を比較する正解率推移を示すグラフである。

【図10】本発明の第5の実施形態の能動学習システムの構成を示すブロック図である。

【図11】図10に示すシステムによる能動学習法と従来の能動学習法とにおける学習精度を比較するROC曲線のグラフである。

【図12】本発明の第6の実施形態の能動学習システムの構成を示すブロック図である。

【図13】図12に示すシステムによる能動学習法と従来の能動学習法とにおける学習精度を比較するグラフであって、(a)はヒット率を示すグラフ、(b)はROC曲線を示すグラフである。

【図14】本発明の第7の実施形態の能動学習システムの構成を示すブロック図である。

【図15】図14に示すシステムによる能動学習法と従来の能動学習法とにおける学習精度を比較するグラフであって、(a)はヒット率を示すグラフ、(b)はROC曲線を示すグラフである。

【図16】本発明の第8の実施形態の能動学習システムの構成を示すブロック図である。

【図17】図16に示すシステムを用いた能動学習法の処理を示すフローチャートである。

【図18】図16に示すシステムによる能動学習法と従来の能動学習法とにおける学習精度を比較する正解率推移を示すグラフである。

【図19】本発明の第9の実施形態の能動学習システムの構成を示すブロック図である。

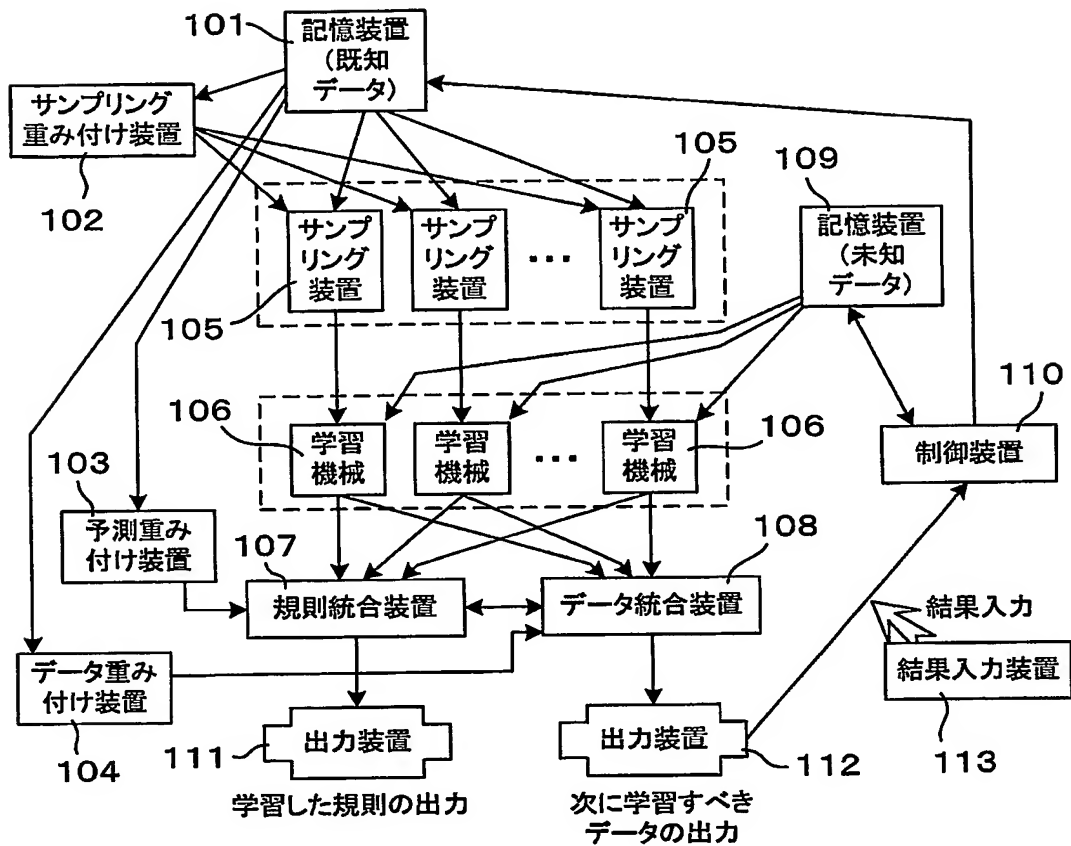
## 【符号の説明】

## 【0073】

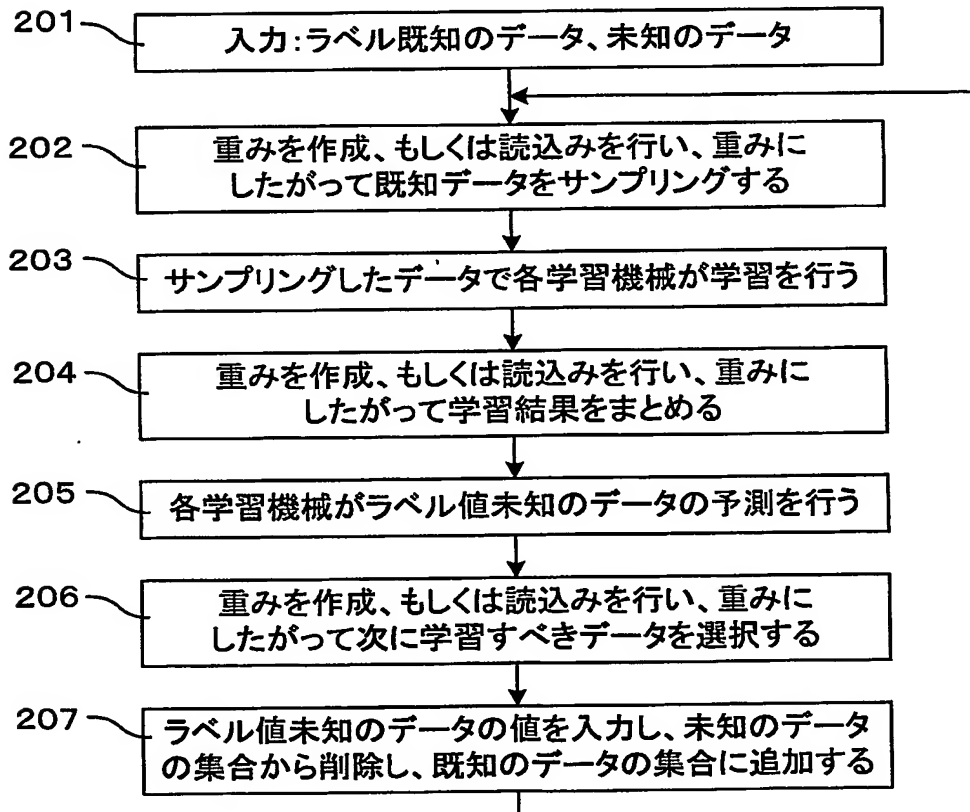
- 101, 109 記憶装置
- 102 サンプリング重み付け装置
- 103 予測重み付け装置
- 104 データ重み付け装置
- 105 サンプリング装置
- 106 学習機械

- 107 規則統合装置
- 108 データ統合装置
- 110 制御装置
- 111, 112 出力装置
- 113 結果入力装置
- 114 データ統合選択装置
- 115 グループ生成装置
- 116 データ選択装置
- 201~207、211~218 ステップ

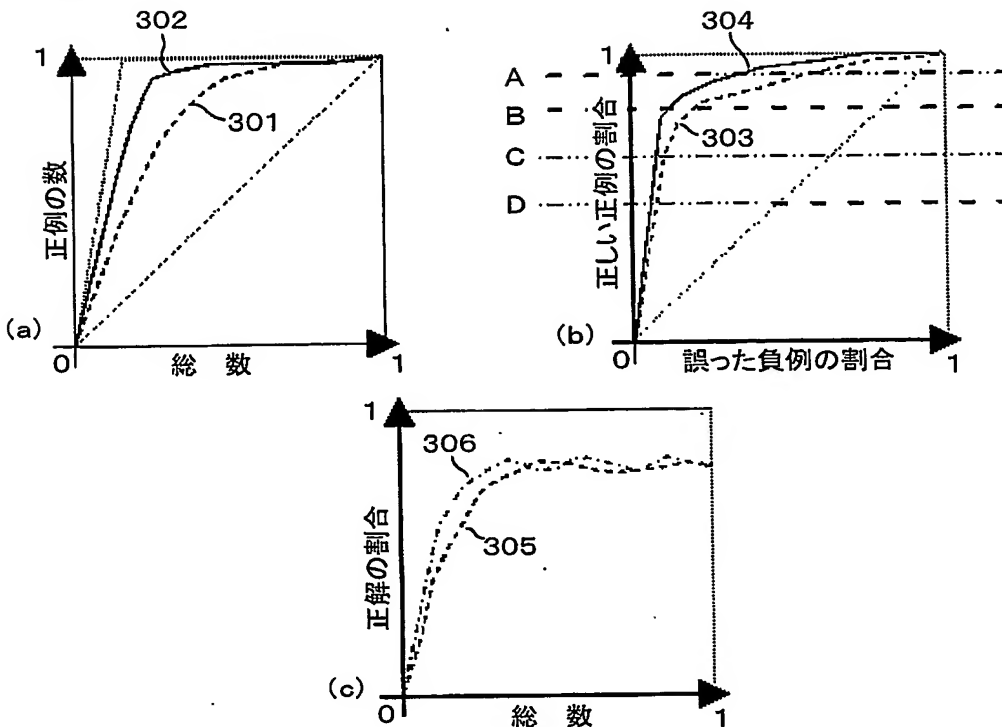
【書類名】 図面  
【図 1】



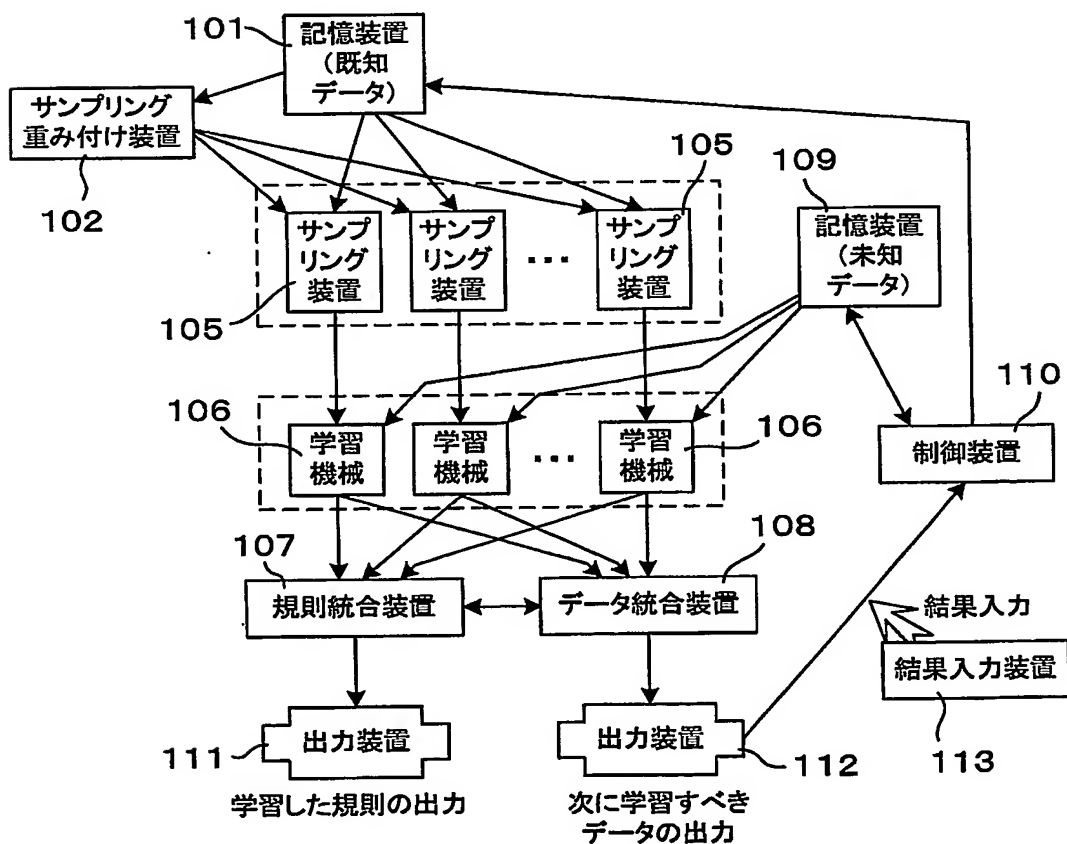
【図 2】



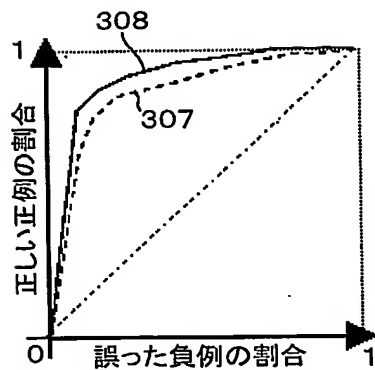
【図 3】



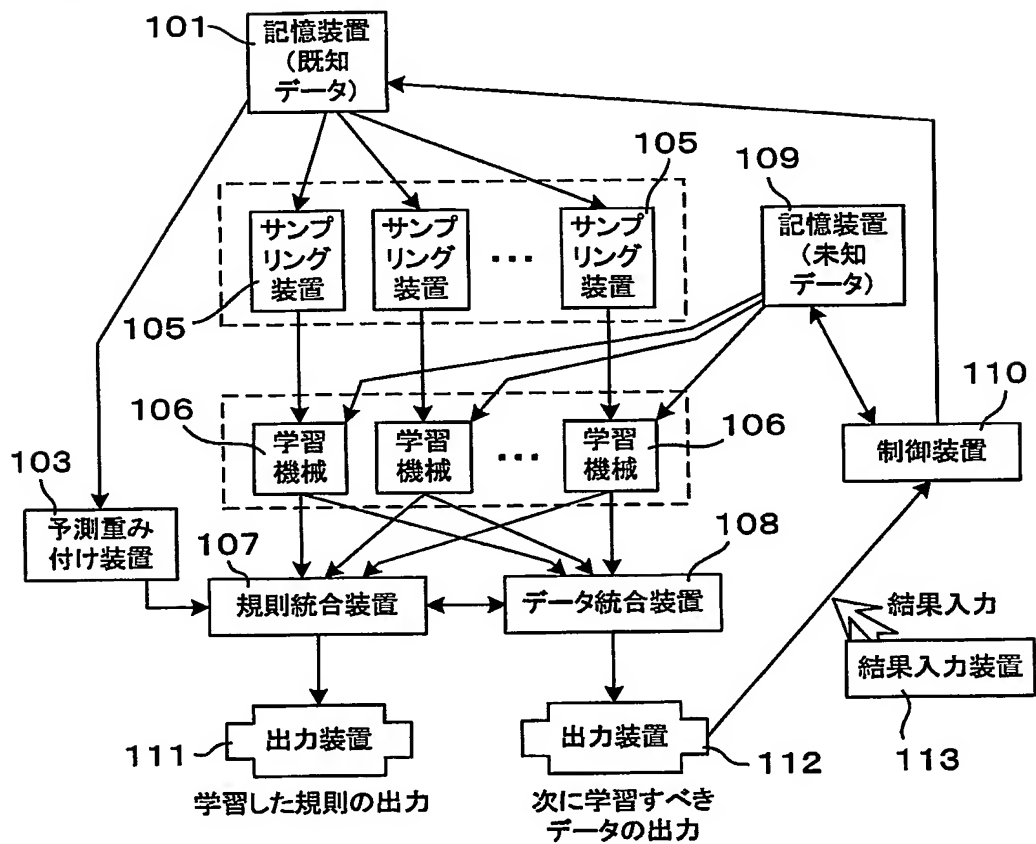
【図 4】



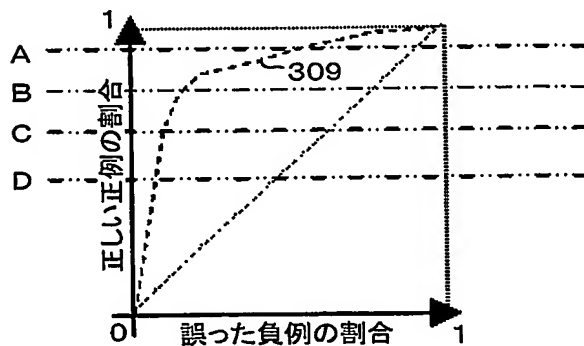
【図 5】



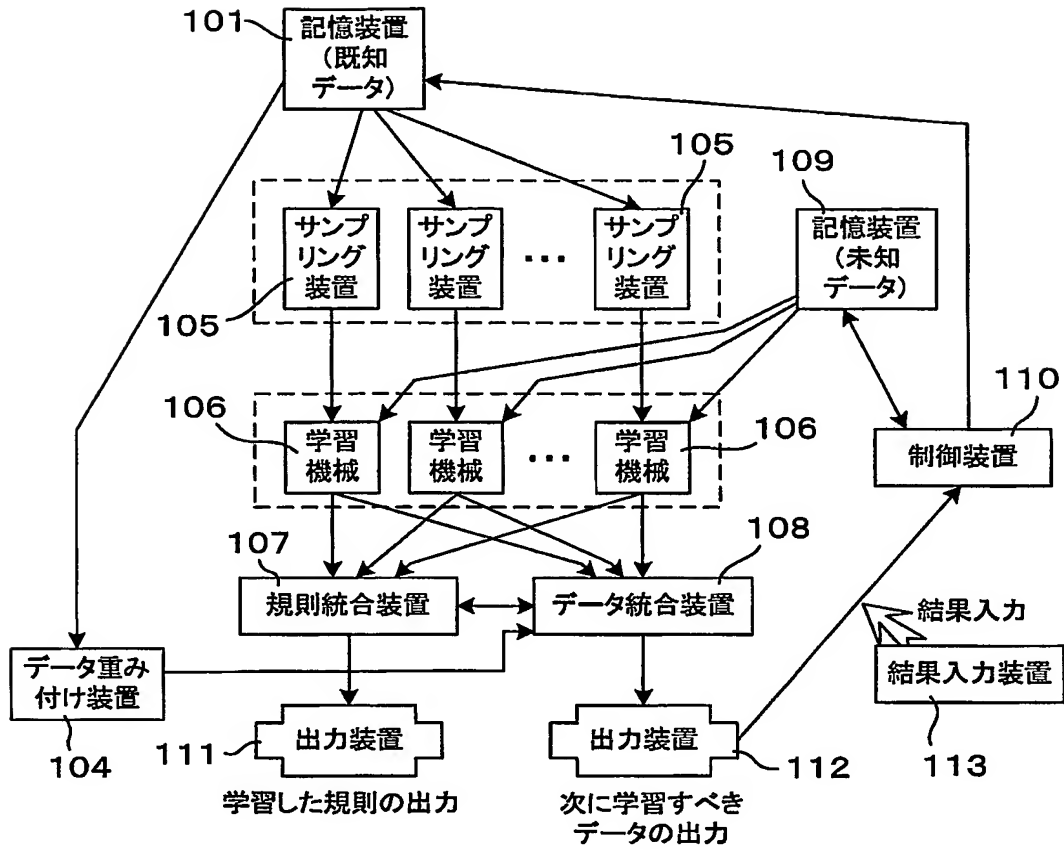
【図 6】



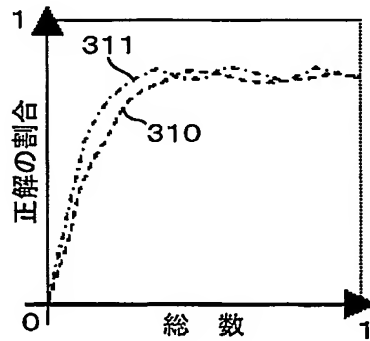
【図 7】



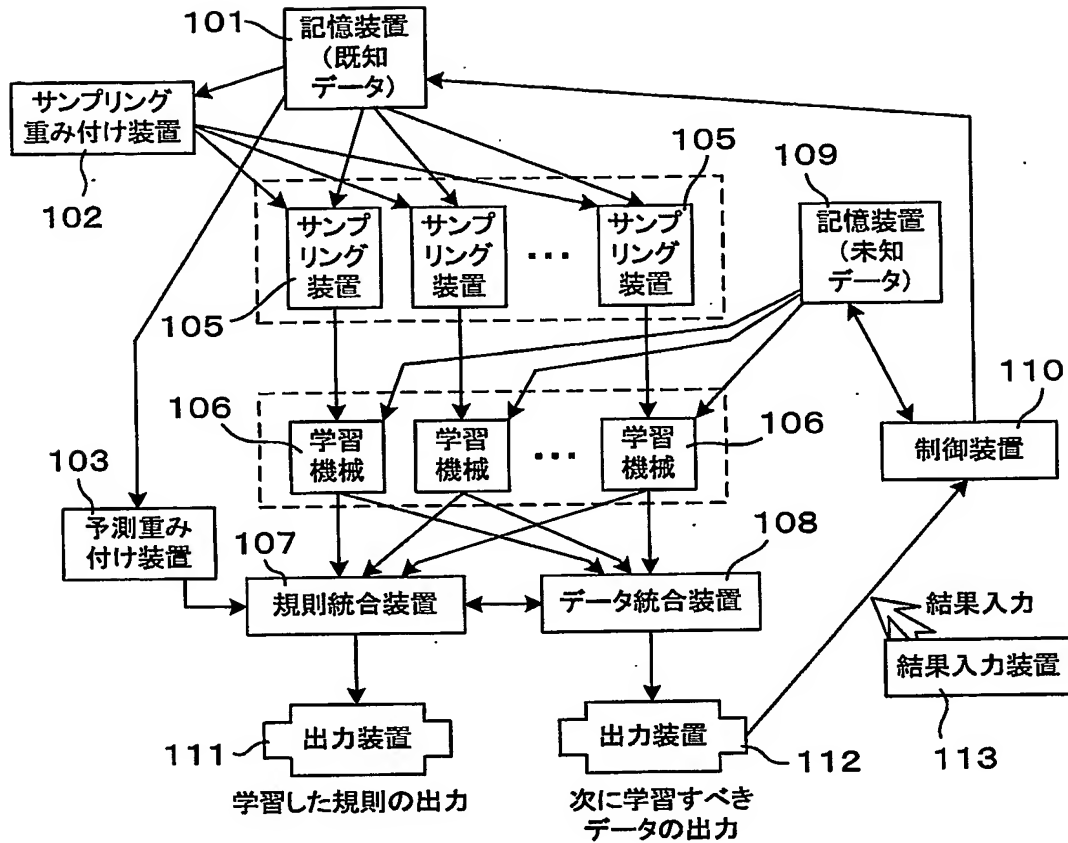
【図 8】



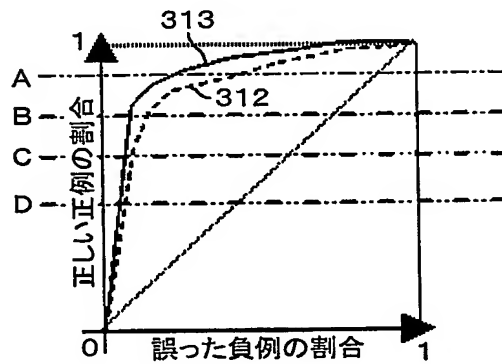
【図 9】



【図10】

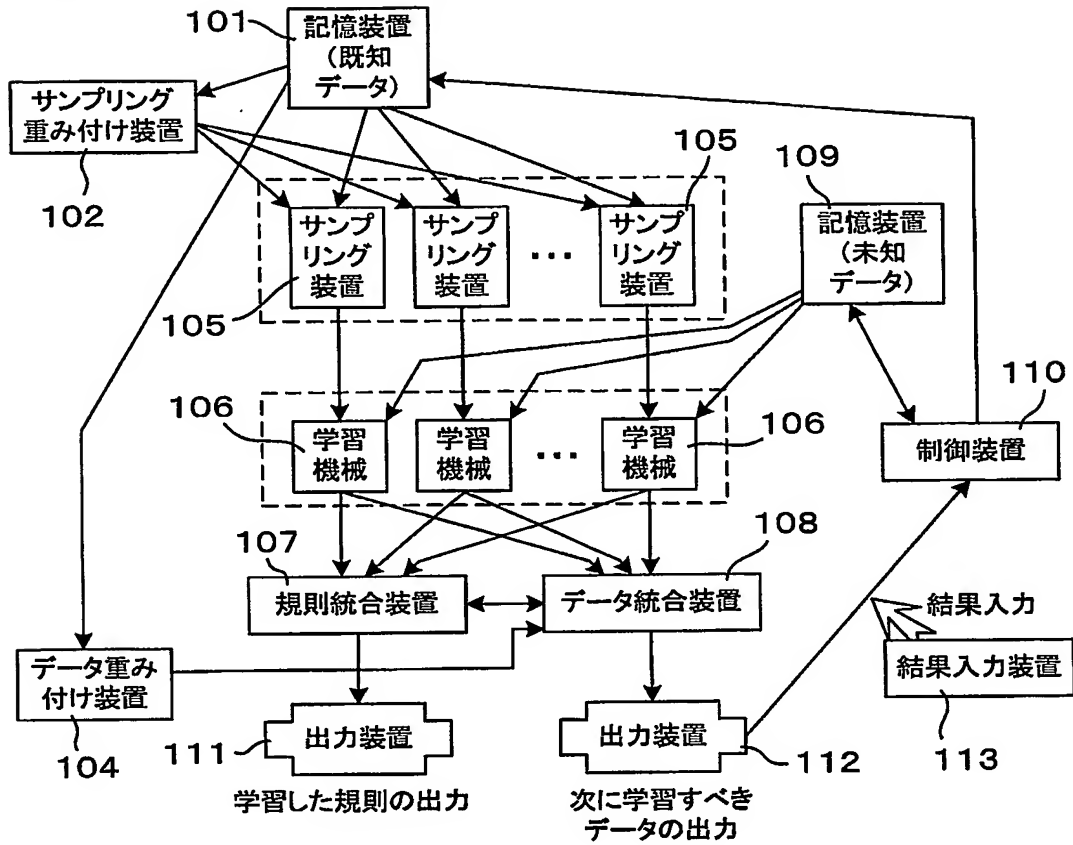


【図11】

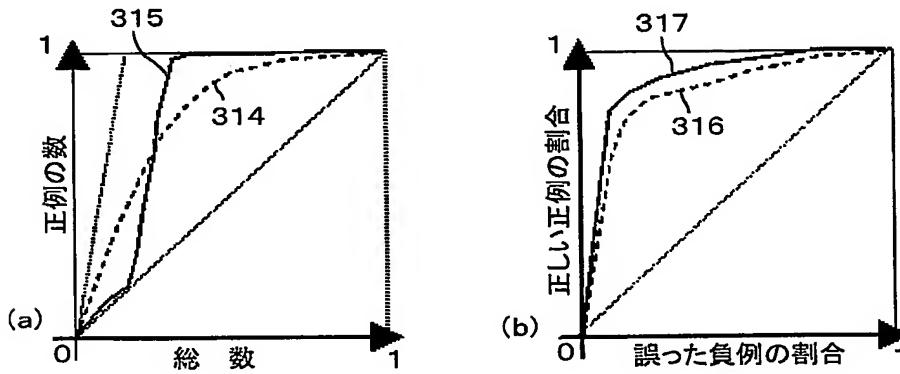




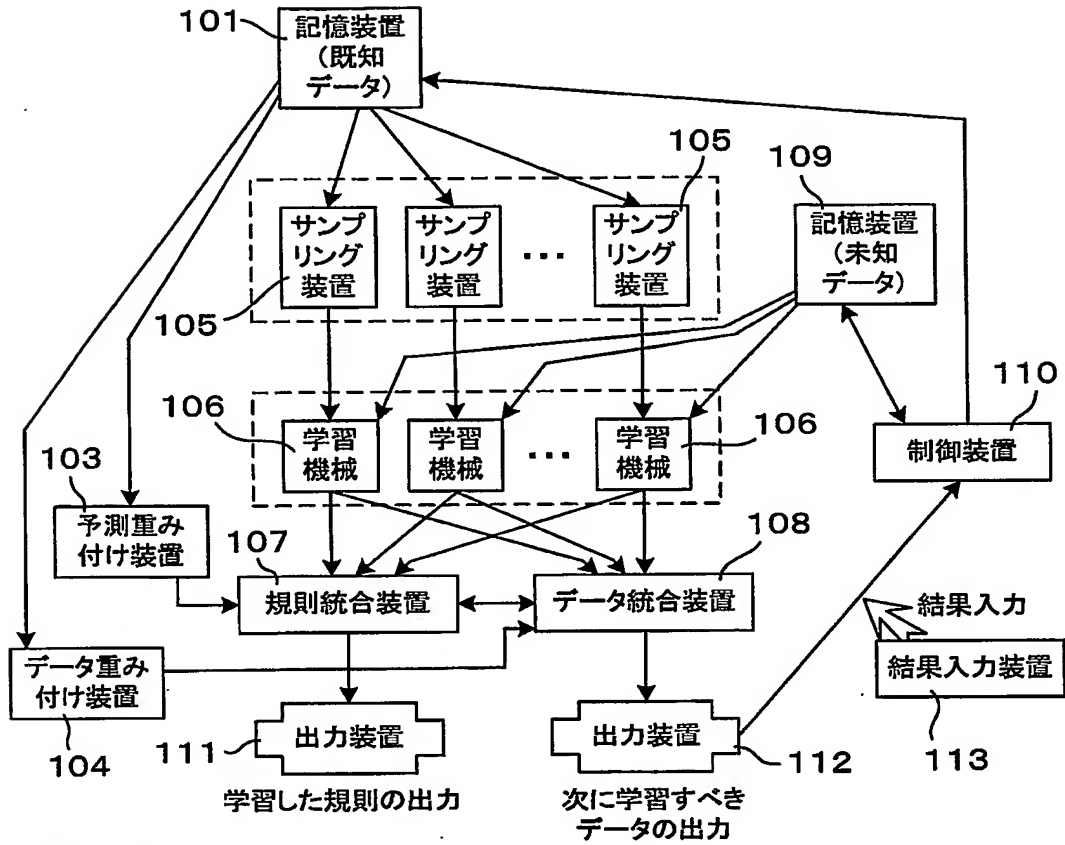
【図 12】



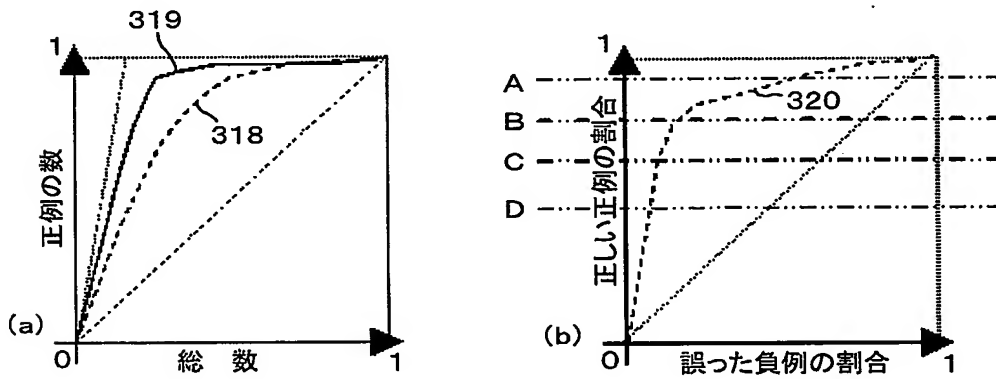
【図 13】



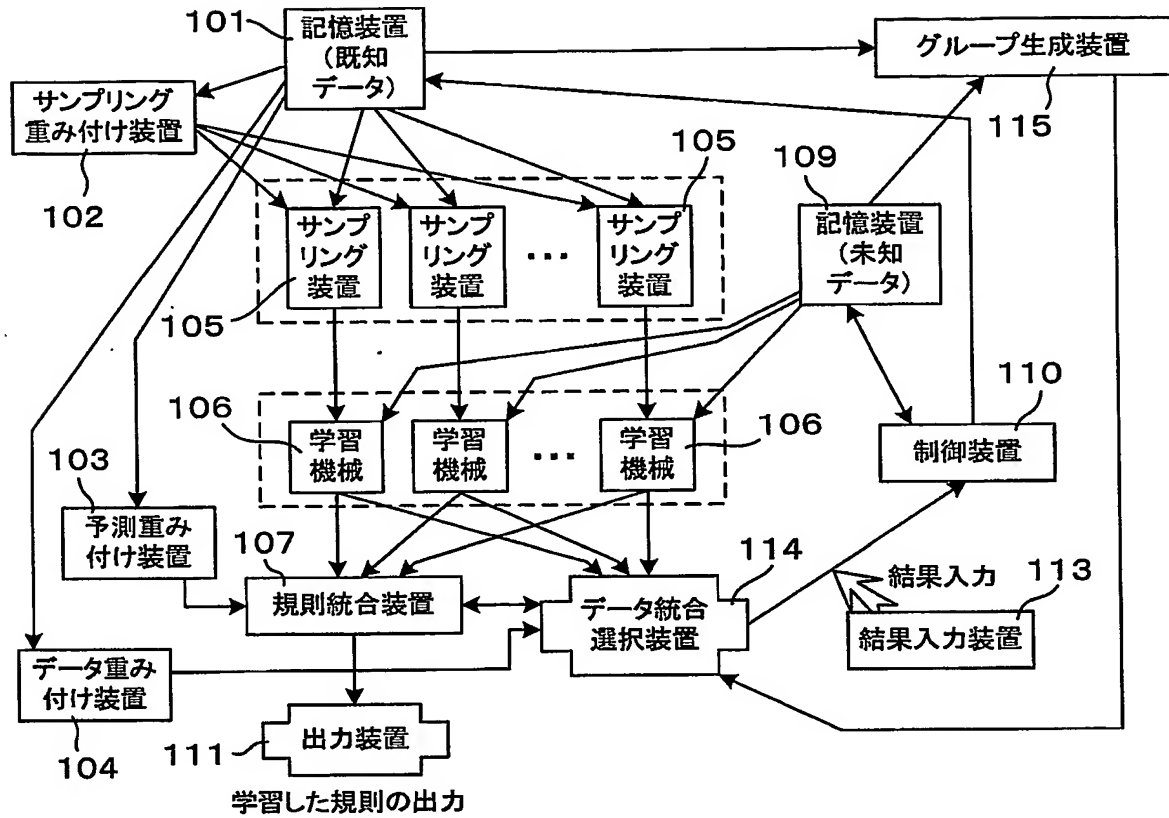
【図 14】



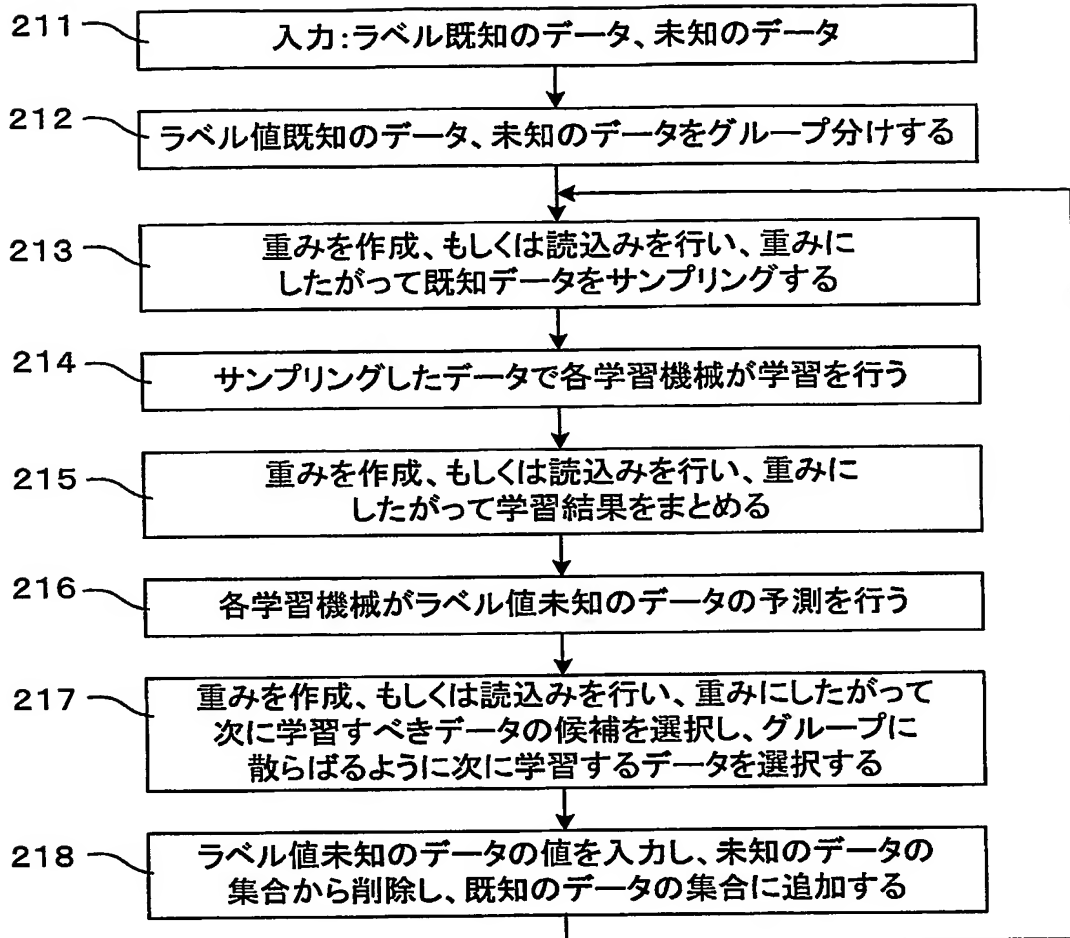
【図 15】



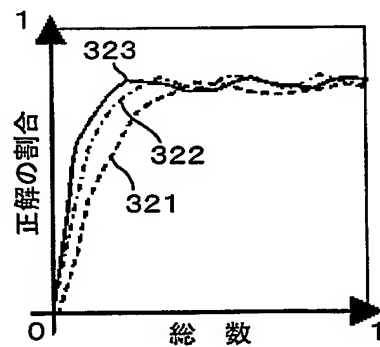
【図16】



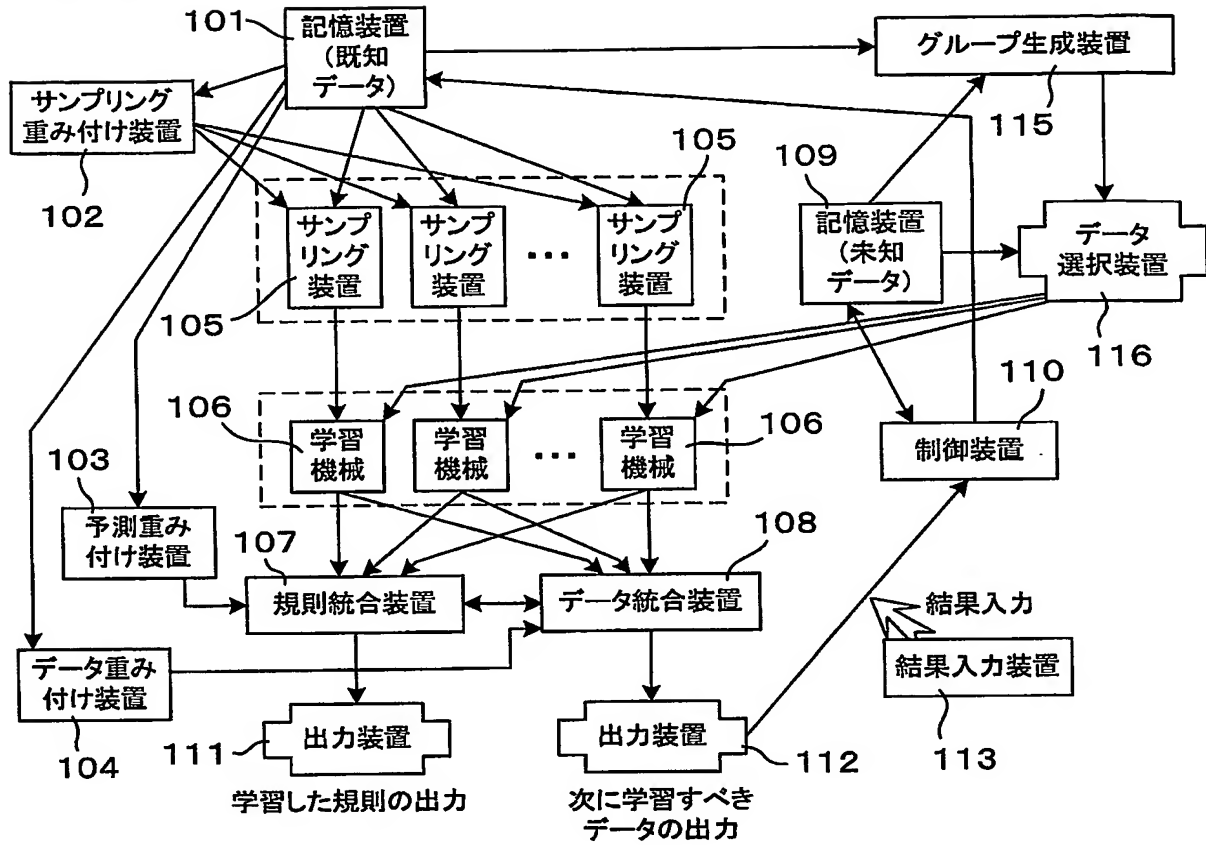
【図 17】



【図 18】



【図 19】



**【書類名】 要約書****【要約】**

**【課題】** 能動学習法の精度を向上させつつ、利用者の意思では制御できなかった精度や、興味のあるデータを先に抜き出すという機能を提供する。

**【解決手段】** 既知データをサンプリングし、既知データに対しては複数の学習機械 106 で独立に学習を行い、未知データに対しては次に学習すべきデータを選択する能動学習システムにおいて、既知データをサンプリングする時に重み付けを行うためのサンプリング重み付け装置 102 と、学習機械 106 の学習結果を統合する際に重み付けを行うため予測重み付け装置 103 と、次に学習すべきデータを選択するときに重み付けを行うためデータ重み付け装置 104 と、を設ける。各重み付け装置においては、データ数に極端に偏りが発生しているときに、その比重を重くするように重み付けを行う。

**【選択図】**

図 1

特願 2003-386702

出願人履歴情報

識別番号

[000004237]

1. 変更年月日

1990年 8月29日

[変更理由]

新規登録

住所

東京都港区芝五丁目7番1号

氏名

日本電気株式会社

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record.**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☒ **BLACK BORDERS**

☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**

☐ **FADED TEXT OR DRAWING**

☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**

☐ **SKEWED/SLANTED IMAGES**

☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**

☐ **GRAY SCALE DOCUMENTS**

☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**

☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**

☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**